

1. はじめに

交通事故を回避するために、運転における危険因子を搭乗者に伝達するシステムの実現が必要とされている。従来のキャプション生成手法 [1] では、一般画像のキャプション生成は可能であるが、走行シーンの危険因子を対象とするキャプションを生成することはできない。そこで本研究では、キャプションの生成に物体検出の結果を Attention として反映させた Neural Baby Talk (NBT) を提案する、これにより注意喚起を目的としたキャプション生成を実現できる。

2. Neural Baby Talk

NBT は、検出した物体を考慮したキャプションを生成する。NBT には物体検出部と言語モデルがあり、入力画像から Region Proposal Network (RPN) により物体検出し、RoI Pooling を施すことで得た特徴量 v_t から、Pointer Network を用いてキャプションに取り入れる物体を決定し、言語モデルにおいては、Adaptive Attention 機構により、生成された単語毎に検出された物体のラベルを適用できる確率を求め、閾値以上である場合に、物体の検出ラベルを考慮したキャプションを生成する。

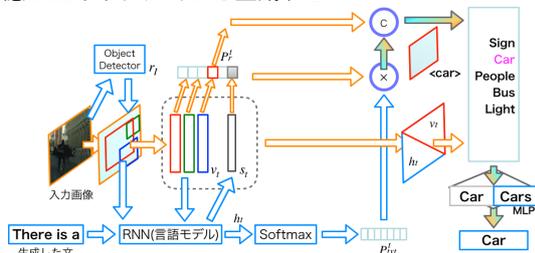


図 1: Neural Baby Talk

3. 提案手法

NBT をベースにした注意喚起に適したキャプション生成法を実現するために、2 つの取り組みを行う。1 つ目は、学習データに対してルールベースの自動アノテーションを施し、注意喚起に適したデータセットを作成する。2 つ目は、NBT に Attention マスクを適用する。これにより、RPN で検出した各領域に注目したキャプションを生成できる。

3.1. ルールベースによる自動アノテーション

走行シーンに対して歩行者や自動車の位置、検出物までの距離を考慮したルールをもとにキャプションのラベルを自動で付与する。アノテーションルールを図 2 に示す。Faster R-CNN [2] により検出した物体のクラス、大きさから基準点 (図 2 の左参照) を算出し、図 2 の右に示す消失点を基準に、方向 (左, 右, 中央) と近距離, 遠距離を考慮したキャプションのアノテーションを作成する。RPN により検出された物体のうち、計 5 種類のクラスを危険因子として定め、歩行者や車、人間が道路上に存在する、信号が存在するなどの危険因子の種類を決定する。

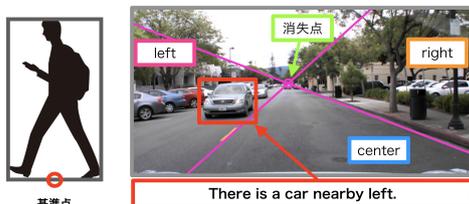


図 2: アノテーションルールの例

3.2. Attention マスクの適用

従来の NBT ではキャプションを生成する際に、検出した全ての領域の特徴量 V を用いて、1 つのキャプションを出力する。しかし、この方法では特定の危険因子に対するキャプションを生成できない。そこで、提案手法では Attention マスク A_t を適用し、図 2 のように特定の危険因子に対する特徴量 v を言語モデルへ入力し、キャプションを生成する。

Attention マスク A_t は検出した危険因子の個数だけ要素を持っており、キャプションに使用する要素を 1 にした

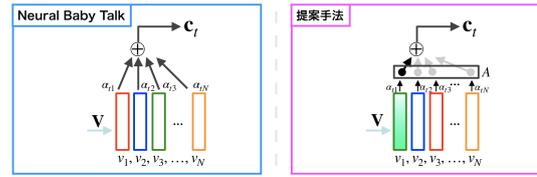


図 3: Attention マスクの適用
one hot vector である。Attention マスクと特徴量 v_i の重み付き和で求めたベクトル c_t は式 (1) で表される。

$$c_t = \sum_{i=1}^N v_i \cdot a_{ti} \cdot A_{ti} \quad (1)$$

4. 評価実験

提案手法の有効性を示すために、評価実験を行う。

4.1. 実験概要

評価方法として、COCO Dataset で学習した NBT と、提案するアノテーションルールと Attention マスクを導入した NBT (NBT+Rule+Att) で比較する。評価には 59 名に対しアンケートを行った。アンケートでは、1 枚の画像に対し従来の NBT および提案手法で生成したキャプションとその画像を表示し、どちらが表示されたシーンに適しているかを選択する。そしてアンケートで得られた適合率と、生成した 69156 キャプション中、検出した物体の語彙が含まれている文の割合である物体関連率で評価する。

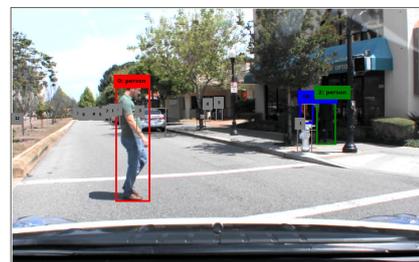
4.2. 実験結果

表 1 に評価結果を示す。評価結果より、アンケート調査では提案手法の適合率が従来手法よりも 20 ポイント優れていることが分かる。提案手法では、左右の方向や、近距離、遠距離の意味が含まれるキャプションが生成されたことから、適合率が上がったと考えられる。また物体関連率に関しても、提案手法は従来手法に比べ、危険因子に注目したキャプションを生成できている割合が高い。

表 1: 評価結果とキャプションの生成例

手法	適合率	物体関連率
NBT	43.1%	29.3%
Our NBT	63.2%	59.3%

図 4 にキャプションの生成例を示す。赤, 緑, 青の順に危険因子の優先度が割り振られており、対応するキャプションの生成例を各色で示す。図 4 より、ルールベースによる危険因子の優先順位の決定により、横断中の人物が最も危険であり、続いて横断の可能性のある人物 2 人を危険因子として判断していることが分かる。また、生成されたキャプションは、それぞれ方向, 距離に対して適切なキャプションが生成することができた。



0. There is a person in front of you on the road.
2. There is a person on the sidewalk in nearby right.
10. There is a person on the sidewalk in front right.

図 4: 注意喚起のためのキャプション生成例

5. おわりに

本研究では、運転シーンにおける注意喚起を目的としたキャプション生成と、その評価実験を行った。今後は複数の危険因子を考慮したキャプションの自動生成を目指す。参考文献

- [1] J. Lu, *et al.*, "Neural Baby Talk", CVPR, 2018.
- [2] S. Ren, *et al.*, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", NIPS, 2015.